Effective number of Hypotheses

- Define: $C[m]$ = maximum number of ways to split m points using concepts in C. (Book calls this $\pi_C(m)$).

- **Theorem:** For any class C, distrib. D, if the number of examples seen m satisfies:

$$m > 2/\epsilon[\log_2(2C[2m]) + \log_2(1/\delta)]$$

then with prob. $(1 - \delta)$, all bad (error $> \epsilon$) hypotheses in C are inconsistent with data.

$C[m]$ is sometimes hard to calculate exactly, but can get a good bound using "VC-dimension". VC-dimension is roughly the point at which C stops looking like it contains all functions.

Shattering

Defn: A set of points S is **shattered** by a concept class C if there are concepts in C that split S in all of the $2^{|S|}$ possible ways.

> In other words, all possible ways of classifying points in S are acheivable using concepts in C.

E.g., any 3 non-collinear points can be shattered by linear threshold functions in 2-D.

VC-dimension

The **VC-dimension** of a concept class C is the size of the largest set of points that can be shattered by C.

> So, if the VC-dimension is d, that means *there exists* a set of d points that can be shattered, but there is *no* set of $d + 1$ points that can be shattered.

E.g., VC-dim(linear threshold fns in 2-D) = 3.

What is the VC dim of intervals on the real line?

How about $C = \{$all boolean functions on n features$\}$?

Upper and lower bound theorems

- **Theorem 1:** $C[m] = O(m^{VCdim(C)})$.

- **Theorem 2:** For any class C, distrib. D, if the number of examples seen m satisfies:

  $$m > 2/\epsilon[\log_2(2C[2m]) + \log_2(1/\delta)]$$

  then with prob. $(1-\delta)$, all bad (error $> \epsilon$) hypotheses in C are inconsistent with data.

- **Theorem 3:** Can replace bound in Theorem 2 with:

  $$\frac{8}{\epsilon}[VCdim(C)\log(1/\epsilon) + \log(1/\delta)]$$

- **Theorem 4:** If $|C| > 2$, then for any learning alg A, there exists a distribution D, and distribution on target concepts in C such that expected error of A is greater than ϵ if A sees less than

  $$\frac{VCdim(C)-1}{8\epsilon}$$

  examples.